Shape-Based Glioma Mutation Prediction Using Magnetic Resonance Imaging

S.J.C. Schielen[A], J.K.H. Spoor[B], R.E.M. Fleischeuer[D], H.B. Verheul[c], S. Leenstra[B], S. Zinger[A]
A: Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands
B: Department of Neurosurgery, Erasmus Medical Center, Rotterdam, the Netherlands
C: Department of Neurosurgery, Elisabeth-TweeSteden hospital, Tilburg, the Netherlands
D: Department of Pathology, Elisabeth-TweeSteden hospital, Tilburg, the Netherlands
Email: s.j.c.schielen@student.tue.nl, j.spoor@erasmusmc.nl, r.fleischeuer@etz.nl, j.verheul@etz.nl, s.leenstra@erasmusmc.nl, s.zinger@tue.nl

***Abstract*—Gliomas are the most frequently occurring primary brain tumors. Determination of the IDH-mutation (Isocitrate DeHydrogenase) in these tumors improves classification and predicts survival. Currently, the only way of determining the mutation status is through a brain biopsy, which is an invasive procedure. This paper concerns the classification of a brain tumor's mutation status through medical imaging. This study proposes a method based on shape description and machine learning. Magnetic resonance images of brain tumors were manually segmented through contour drawing, then analyzed through mathematical shape description. The extracted features were classified using multiple algorithms of which Random Undersampling Boosted Trees gave the highest accuracy. An accuracy of 86.4% was found using leave-one-out cross-validation on a data set of 13 IDH-positive and 9 IDH-wild-type gliomas. The results indicate the feasibility of the proposed approach, but further research on a larger data set is required.**

I. Introduction

Gliomas are the most frequently occurring primary brain tumors. Gliomas are graded based on their malignancy. They are classified as a grade I-IV (low abnormality - high abnormality) on the World Health Organization (WHO) scale [1]. The most important molecular marker that influences the prognosis is the tumor's IDH-status (Isocitrate DeHydrogenase). A glioma's IDH-status is either positive to a mutation (IDHpos) or not mutated: the wild type (IDHwt). The IDH-status is vital to know, as it substantially influences the treatment prognosis. It is common for gliomas to be visualized through Magnetic Resonance Imaging (MRI).

In conventional MRI, typically different modalities are used. These modalities are: T1-weighted contrast-enhanced (T1CE) and Fluid Attenuated Inversion Recovery (FLAIR) among others.

Despite efforts by experienced doctors with a trained eye, distinguishing a glioma's IDH-status by visual inspection of MRI data does not give certainty. Therefore, current methods only allow the IDH-status to be identified through a biopsy. A biopsy requires invasive brain surgery, which is accompanied by surgery related risks and discomfort for the patient.

This research aims to explore the feasibility of brain tumor shape analysis on conventional MR images to substitute a biopsy for detecting the IDH-status of gliomas. This paper has two main contributions. Two new metrics are proposed and successfully applied to solve this problem. The novelty of the metrics consists of including image-based context information, such as relative tumor size and perimeter, and the tumor's location as input to the computer-aided decision support system. The second contribution consists of obtaining the first experimental results on a clinical data set consisting of gliomas with and without a mutation. An accuracy of 86.4% is found. This shows the potential applicability of the proposed methodology in clinical practice.

A state-of-the-art research review is given in Section II. Section III contains the methodology used in this study. Then, the results are listed in Section IV and the drawn conclusions are given in Section V.

II. State of the Art

The general approach to medical image analysis consists of the following steps: image preprocessing, feature extraction and classification. Image preprocessing is done to enhance the informational quality of the images. MR images are, as all medical imaging systems, prone to noise. There are several techniques to filter out noise, but this usually comes at the cost of losing information that may be valuable to the classifier [2]. Although conventional segmentation algorithms exist to segment images, applying these to an MR image of a brain poses difficulties. Chen et al. [2] propose a two-dimensional level set algorithm to create a contour around a breast tumor, which had satisfactory results. Gliomas, however, have less contrast between the tumor and adjacent tissues than breast tumors. Therefore, manual contours are a time consuming yet effective method of feature extraction for gliomas, as also performed by Jeong et al. [3]. Moreover, Limkin et al. [4] state that brain tumors can have 'ill-defined borders', which not only adds to the difficulty of segmentation, it also indicates that the tumor may spread to adjacent tissues. How vaguely defined the edges of a tumor are on the image can be an indication of malignancy. It may be a direct indication of the tumor's IDH-status. Although it may be supported by experience, manual contouring is prone to subjectivity.

The contoured tumors are used for feature description. Zacharaki et al. [5] used a combination of two-dimensional

(2D) tumor shape characteristics, patient data such as age, and image intensity data within the regions of interest (ROI). Besides medical images, many factors such as age, gender and lifestyle are interwoven in the occurrences of tumors. As there is a link between gliomas and patient metadata, 2D shape descriptors combined with patient metadata may outline the valuable information contained in the extracted features.

Shape descriptors, both applied on 2D and 3D shapes, are used in feature description applied on gliomas. Jeong et al. [3] used a combination of 3D shape and textural features, which yielded an accuracy of 96% in differentiating low grade from high grade gliomas. Zacharaki et al. [5] found an accuracy of 88%, using 2D shape characteristics, texture characteristics and external characteristics. Although the outcomes of these studies show relatively high accuracies, it is unknown how much only shape features or texture features contribute individually to the classification of the IDH-status.

Shape description can be oriented toward contours and regions: the edge of and area filling the shape, respectively [6]. Some contour-oriented descriptors are perimeter, circularity measures such as normalized perimeter to area [7] and Haralick's Circularity Measure [8], compactness measures such as Danielson's Shape Factor [9] and Bribiesca's Normalized Discrete Compactness [10], Grandlund's Fourier Descriptors [11], eccentricity and contour sequence moment coefficients [12]. Some region-oriented descriptors are area, centroid, area ratio [13] and mean radial length. Since digital images consist of pixels, there tends to be an error when applying these shape descriptors with respect to continuous shapes [7].

Which machine learning algorithm gives the highest accuracy as classifier depends on the specific features. Some reported cases (all applied to MR images of gliomas) found the highest accuracies by using Decision Trees (DT) in Random Forests (RF) [3], [14], [15]. Other reported cases found the highest accuracies by using a Support Vector Machine (SVM) [5], [16]. Although RFs and SVMs are frequently used, the feature values are different in each case. Therefore, which classifier performs best cannot be determined a priori, as described by Sekeroglu et al. [17]. What all of the aforementioned cases agree on is testing out different algorithms and assessing them on their accuracy, sensitivity and specificity. As for the validation, most cases used k-fold cross-validation [14]. In the case of smaller sample sizes, the leave-one-out cross-validation is used [3], [5].

III. Methodology

To predict an IDH-mutation for gliomas without a biopsy, the proposed approach is shown step by step in Figure 1. The system takes a contoured MR image as input and provides a binary prediction as output.

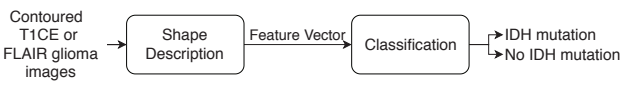

Fig. 1. Flowchart for MR image processing

A. Shape Description

The motivation behind the use of contours lies in the simplicity and low computational cost of the feature extraction with respect to texture analysis. Moreover, contours are independent of tumor biology which is quite complex and influences the texture.

Multiple mathematical shape descriptions of the tumor's contour are required to obtain the useful information needed to express the tumor as an array of features. The respective differences between tumors' features are hypothesized to distinguish a mutated glioma from a wild-type glioma.

Each of the contours was saved as a logical matrix, in which ones represented the contour pixels. These pixels were described by an array of coordinates. The contour is considered as a non-self-intersecting polygon with N nodes (pixels) in order: $(x_0, y_0), (x_1, y_1), ..., (x_{N-1}, y_{N-1})$. Since the contours are closed, the N-th node is the same as the first node: $(x_N, y_N) = (x_0, y_0)$. The following shape descriptors were used.

The perimeter P was computed by

$$P = \sum_{n=0}^{N-1} \sqrt{(x_{n+1} - x_n)^2 + (y_{n+1} - y_n)^2}, \tag{1}$$

hence P is a sum of Euclidean distances. The area A was computed by Gauss's area formula:

$$A = \frac{1}{2} | \sum_{n=0}^{N-1} (x_n y_{n+1} - x_{n+1} y_n) |. \tag{2}$$

The centroid coordinates C_x and C_y were calculated by

$$C_x = \frac{1}{6A} \sum_{n=0}^{N-1} (x_n + x_{n+1})(x_n y_{n+1} - x_{n+1} y_n), \tag{3}$$

$$C_y = \frac{1}{6A} \sum_{n=0}^{N-1} (y_n + y_{n+1})(x_n y_{n+1} - x_{n+1} y_n), \tag{4}$$

respectively.

The radial length RL of contour point n is the distance from (x_n, y_n) to the centroid:

$$RL[n] = \sqrt{(x_n - C_x)^2 + (y_n - C_y)^2}. \tag{5}$$

This was normalized through

$$NRL[n] = \frac{RL[n]}{max(RL[n])}, \tag{6}$$

where $NRL[n]$ is the normalized radial length to contour pixel n. Computing the normalized radial length for all N contour pixels yields an array of numbers, so the mean of this was taken as a single feature:

$$\mu_{NRL} = \frac{1}{N} \sum_{n=0}^{N-1} NRL[n], \tag{7}$$

where μ_{NRL} is the mean normalized radial length. The eccentricity ϵ is a unique measure of how much a shape deviates from a circle. It can be calculated by

$$\epsilon = \frac{(\mu_{2,0} - \mu_{0,2})^2 - 4(\mu_{1,1})^2}{(\mu_{2,0} + \mu_{0,2})^2}, \tag{8}$$

where $\mu_{2,0}$, $\mu_{0,2}$ and $\mu_{1,1}$ are second order central moments.

The following shape descriptors were also computed, because they have been useful in the researched studies. The roughness index RI [13], the area ratio AR [13], normalized perimeter to area $NP2A$ [7], Haralick's Circularity Measure HC [8], Danielson's Shape Factor G [9], Bribiesca's Normalized Discrete Compactness C_{dn} [10], the convexity $conv$ [18], Grandlund's Fourier Descriptors $\bar{FD}$ [11] and contour sequence based shape descriptors $F1$, $F2$, $F3$, $F4$ [12].

B. Clinically Inspired Shape Descriptors

Since there were variations in the MR images' dimensions and the amount of black pixels surrounding the head in the image, normalization was necessary to ensure a correct comparison between tumors. A binarized image of the head was created through thresholding, where the threshold was determined manually for each image. By subsequent erosion and dilation, a closed shape of the head was found. By using the Sobel edge detection algorithm, the outline of the head was found. Since these edges were usually quite rough, a convex hull around the head's outline was used to smoothen the contour. This slightly increases the area inside of the head contour, but substantially lowers the perimeter. Only the parameters P, A, C_x and C_y had to be normalized with respect to the head, since the other parameters were normalized already, invariant to the size or a ratio of P and A. The area of the tumor was normalized, with respect to the head, by

$$A_n = \frac{A_t}{A_h}, \tag{9}$$

where A_n is the normalized area, A_t is the tumor's area and A_h is the head's respective area. The normalized perimeter P_n can be calculated similarly to equation (9).

The tumor's centroid (C_{tx}, C_{ty}) was normalized with respect to the head's centroid (C_{hx}, C_{hy}) to obtain an indication of location. The normalized centroid (C_{nx}, C_{ny}) was computed by

$$C_{nx} = \frac{C_{tx} - C_{hx}}{\frac{1}{2} l_{head}}, \; C_{ny} = \frac{C_{ty} - C_{hy}}{\frac{1}{2} w_{head}}, \tag{10}$$

where l_{head} and w_{head} are the respective length and width of the head's contour. The division by l_{head} and w_{head} was done to normalize and the factor $\frac{1}{2}$ positions the head's centroid at the origin when C_{nx} and C_{ny} are plotted. A plot of the coordinate vectors of both the tumor's contour and head's contour along with the respective centroids is shown in Figure 2. This is a representation of the same contoured tumor shown in Figure 3. Since the resolutions of the images were different, the normalization is vital to indicate position. Otherwise, the same tumor could have different centroid coordinates based on the resolution of the image. Normalization ensures a correct comparison between images of different resolutions.

The feature vector resulting from shape description consists of the following 21 shape descriptors: AR, A_n, C_{dn}, $F1$, $F2$, $F3$, $F4$, the second until the fifth Fourier Descriptors $\bar{FD}[2-5]$, G, HC, $NP2A$, P_n, RI, $conv$, ϵ, μ_{NRL}, C_{nx} and C_{ny}. The feature vector was computed for each of the contours and then used as input to the classification algorithms.

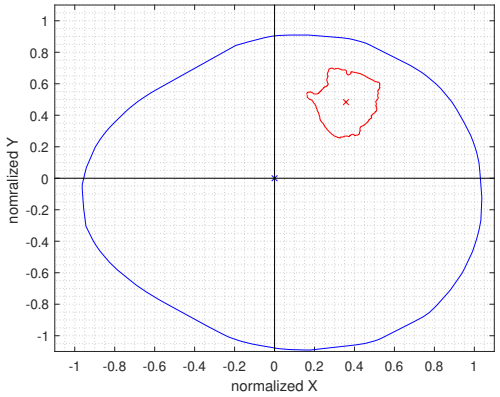

Fig. 2. The head contour (blue), normalized head centroid (blue cross at origin), tumor contour (red) and tumor centroid (red cross) of an IDH mutated glioma

C. Classification

The extracted features were labelled with the tumor's IDH-status obtained from tissue analysis to allow supervised learning. Several classification algorithms were applied such as a Decision Tree (DT), Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). These classifiers were applied in multiple degrees of complexity and also in ensemble methods. All these algorithms were trained with the data set and then assessed.

To train the classifier as much as possible with the relatively small data set, the leave-one-out cross-validation was used. The leave-one-out cross-validation computes the prediction accuracy by iteratively training on all samples but one and testing on the left out sample. This is repeated until all samples have been the test sample once. The final accuracy is the average of each iteration's accuracy. Furthermore, the sensitivity and specificity were computed, which are respectively also known as the True Positive Rate (TPR) and True Negative Rate (TNR). Through the TPR and TNR a Receiver Operator Characteristic (ROC) curve is created, which allows the Area Under the Curve (AUC) to be computed. The classifiers were assessed based on their accuracy, AUC, TPR and TNR.

The data set was rather small, as 22 patients were used whereas researched studies used at least 40. The rather small data set increased the risk of a classifier that is not general enough; overfitted. Since the data set consisted of both FLAIR and T1CE images, there were two types of images for most of the patients. Despite the normalization efforts, there were (small) differences in the shape descriptor values between the same tumor on a FLAIR image and on a T1CE image, which was also due to the manual contour delineation. To test for overfitting, the classifier was used to predict the same tumors on the FLAIR images. The validation accuracy (VA) is the accuracy obtained by training and testing using leave-one-out cross-validation on T1CE images. The test accuracy (TA) is the accuracy found when the trained classifier was used on FLAIR images of the same tumors that were used to train on. Since the differences between the pictures should be small,

TABLE I
PERFORMANCE OF THE BEST CLASSIFIERS

Classifier	VA	TA	TPR	TNR	AUC
DT	77.3%	50%	76.9%	77.8%	0.78
Quadratic SVM	68.2%	81.25%	84.6%	44.4%	0.58
RUSB Trees	86.4%	87.50 %	84.6%	88.9%	0.86
SUB KNN	72.7%	75%	84.6%	55.6%	0.66

it was expected that the TA was at least the VA. If it was significantly lower, the classifier was probably overfitted.

IV. RESULTS

In this section, the used data set and tools that were found to be useful are mentioned. The results of applying the methodology are listed and discussed.

A. Data Set

The data set used in this study consists of 22 patients. Both FLAIR and T1CE images were present for most of the patients. For some of the patients, only T1CE images were available. The available images were in the form of a 3D matrix, which represents the patient's entire brain. Each image is one layer of the brain, which is also referred to as a slice. The resolutions of the images differ per patient: they range from 250 by 250 pixels to 1100 by 1100 pixels. The set consisted of 13 IDHpos and 9 IDHwt cases confirmed through immunohistochemistry. There are more IDHpos cases than IDHwt cases in the data set, therefore the data set is considered imbalanced. Since there were no contoured images available, this was done manually. Both the FLAIR and T1CE images were contoured.

Figure 3 shows a manually contoured tumor. This was made using the MATLAB Image Processing Toolbox. The function 'drawassisted' allows the placement of way points around the ROI. These way points are connected by lines that have a built-in edge detection algorithm. Once the contour is closed, a matrix filled with zeros of the same size as the input image is saved, in which the ROI is filled with ones. By using the Sobel edge detection algorithm, the contour can be extracted.

The exact edges of a tumor may be hard to define due to a number of possible factors. The resolution may be low around the tumor. The tissues surrounding the tumor may be irritated due the exerted pressure by the tumor and therefore appear as tumor tissue on the MR image. Moreover, the manual delineation of a tumor is subjective. Therefore, the contours are validated by a neurosurgeon.

B. Classification

The classification was performed using Matlab's application Classification Learner. The results of the best four classifiers are listed in Table 1. The TPR, TNR and AUC are computed for the leave-one-out cross-validation within the application. It should be noted that the DT has a higher validation accuracy than test accuracy. Since its test accuracy is 50%, it is the same as a random binary classifier like flipping a coin. Therefore, this classifier is most likely overfitted to the training data.

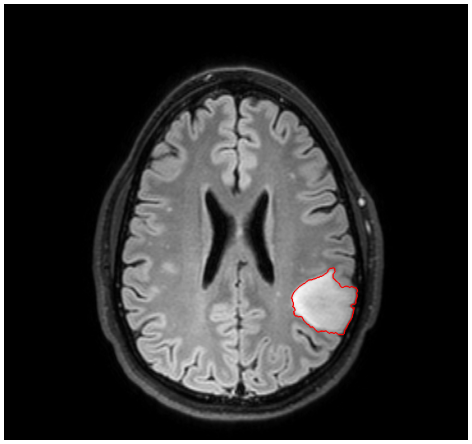

Fig. 3. A manually contoured tumor on a FLAIR image of an IDH mutated glioma

The best performance is found by using Random Undersampling Boosted Trees (RUSB Trees). The confusion matrix of this classifier is listed in Table II and the ROC is shown in Figure 4.

TABLE II
THE CONFUSION MATRIX OF THE RUSB TREES CLASSIFIER ON A DATA SET OF 9 IDHWT AND 13 IDHPOS GLIOMAS

		Predicted class IDHpos	Predicted class IDHwt
True class	IDHpos	11	2
True class	IDHwt	1	8

RUSB Trees [19] is an ensemble classifier in which multiple trees are used. The accuracy is high, because RUSB Trees uses the concepts of undersampling and boosting. Random undersampling is used to lower the observations of the larger set in the whole set of observations to the same size of the smaller set. This does come at the cost of losing some information, but it balances the observations, making overfitting to the frequently occurring observations less likely. Boosting is an iterative procedure in which weights are modified depending on the performance. After the last iteration, all trees get a vote on the newly obtained observation (test data), in which the weight determines how much the vote contributes. The highest sum of these weighted votes is the classification.

Since there is a random aspect in what observations are left out, the accuracy of RUSB Trees can fluctuate each time the classifier is trained. This implies that there still is a chance of overfitting, depending on how the data points that are used to train the classifier represent the general trend the mutations of these gliomas tend to have. In some runs an accuracy of 90% was found, accompanied by a test accuracy of 50%.

V. CONCLUSIONS

A glioma's IDH-mutation status is vital to know, which currently can only be determined through a biopsy. To avoid

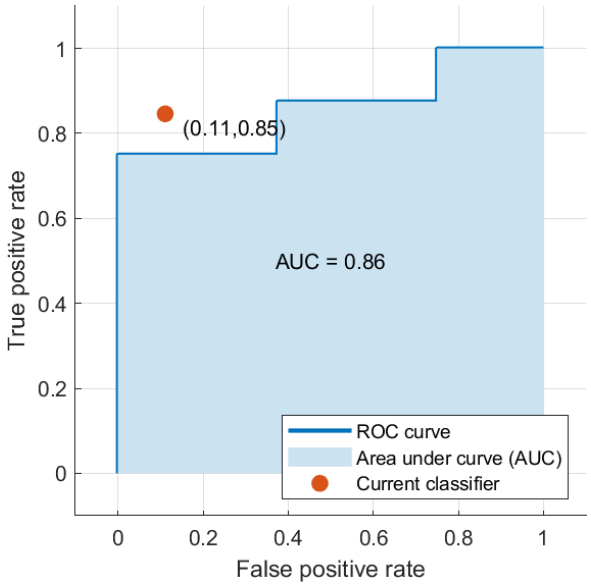

Fig. 4. The ROC of the RUSB Trees classifier on a data set of 9 not mutated and 13 mutated gliomas. The 'current classifier' dot shows the false positive rate and true positive rate of the predictions on the whole data set.

this invasive procedure, a method based on shape analysis of MRI data was proposed.

This study was conducted to assess the feasibility of using MR images to determine whether a glioma is mutated. The proposed method consists of extracting features through shape description and classifying them using machine learning. A new metric was introduced: including image based context information to normalize area, perimeter and indicate position. Another contribution was made by obtaining experimental results on a clinical data set of gliomas with and without mutations.

The best results were found using Random Undersampling Boosted Trees, with a validation accuracy of 86.4%, test accuracy of 87.50%, sensitivity of 84.6%, specificity of 88.9% and AUC of 0.86. Therefore, the results indicate that a method based on the shape description of gliomas in MR images is feasible.

Although the results indicate feasibility, the size of the data set limits certainty. It is unknown how well the observations used in this study represent the general trend of gliomas. In future research, a study with a larger data set should be conducted to give a more verified conclusion. There is no indication that the same classifier that gave the highest accuracy in this study, will yield the highest accuracy in future studies. Therefore, it is advised to use multiple classification algorithms and assess them based on their accuracy, sensitivity and specificity. Although shape description had satisfactory results in this study, combining shape features with texture features is of interest in future work.

It may also be of interest to add patient metadata to the feature vector. As former research indicates the affiliation between gliomas and patient data such as age and gender, valuable information may be left out of the classifier by only considering medical images.

References

[1] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," in *Acta Neuropathologica*, vol. 131, pp. 803–820, 2016.

[2] D. R. Chen, Y. W. Chang, H. K. Wu, W. C. Shia, and Y. L. Huang, "Multiview Contouring for Breast Tumor on Magnetic Resonance Imaging," *Journal of Digital Imaging*, pp. 713–727, 2019.

[3] J. Jeong, L. Wang, B. Ji, Y. Lei, A. Ali, T. Liu, W. J. Curran, H. Mao, and X. Yang, "Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images," *Quantitative Imaging in Medicine and Surgery*, vol. 9, no. 7, pp. 1201–1213, 2019.

[4] E. J. Limkin, S. Reuzé, A. Carré, R. Sun, A. Schernberg, A. Alexis, E. Deutsch, C. Ferté, and C. Robert, "The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features," *Scientific Reports*, vol. 9, no. 1, 2019.

[5] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.

[6] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.

[7] R. S. Montero and E. Bribiesca, "State of the art of compactness and circularity measures," in *International Mathematical Forum*, vol. 4, pp. 1305 – 1335, 2009.

[8] R. M. Haralick, "A Measure for Circularity of Digital Figures," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-4, no. 4, pp. 394–396, 1974.

[9] P. E. Danielson, "A new shape factor," *Computer Graphics and Image Processing*, vol. 7, no. 2, pp. 292–299, 1978.

[10] E. Bribiesca, "Measuring 2-D shape compactness using the contact perimeter," *Computers and Mathematics with Applications*, vol. 33, no. 11, pp. 1–9, 1997.

[11] G. H. Granlund, "Fourier Preprocessing for Hand Print Character Recognition," *IEEE Transactions on Computers*, vol. C-21, no. 2, pp. 195–201, 1972.

[12] L. Gupta and M. D. Srinath, "Contour sequence moments for the classification of closed planar shapes," *Pattern Recognition*, vol. 20, no. 3, pp. 267–272, 1987.

[13] A. V. Alvarenga, W. C. Pereira, A. F. C. Infantosi, and C. M. De Azevedo, "Classification of breast tumours on ultrasound images using morphometric parameters," in *2005 IEEE International Workshop on Intelligent Signal Processing - Proceedings*, pp. 206–210, 2005.

[14] J. Amin, M. Sharif, M. Raza, and M. Yasmin, "Detection of Brain Tumor based on Features Fusion and Machine Learning," in *Journal of Ambient Intelligence and Humanized Computing*, 2018.

[15] T. T. Tang, J. A. Zawaski, M. W. Gaber, K. N. Francis, and A. A. Qutub, "Image-based Classification of Tumor Type and Growth Rate using Machine Learning : a preclinical study," no. July 2018, pp. 1–10, 2019.

[16] A. Vamvakas, S. C. Williams, K. Theodorou, E. Kapsalaki, K. Fountas, C. Kappas, K. Vassiou, and I. Tsougos, "Imaging biomarker analysis of advanced multiparametric MRI for glioma grading," *Physica Medica*, vol. 60, pp. 188–198, 2019.

[17] B. Sekeroglu, S. S. Hasan, and S. M. Abdullah, "Comparison of Machine Learning Algorithms for Classification Problems," in *Advances in Intelligent Systems and Computing*, vol. 944, pp. 491–499, 2020.

[18] A. Boujelben, A. C. Chaabani, H. Tmar, and M. Abid, "Feature extraction from contours shape for tumor analyzing in mammographic images," in *DICTA 2009 - Digital Image Computing: Techniques and Applications*, pp. 395–399, 2009.

[19] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," in *Proceedings - International Conference on Pattern Recognition*, 2008.